Linear Regression (Excel) and Cellular Respiration for Biology, Chemistry and Mathematics

Irene Corriette, Beatriz Gonzalez, Daniela Kitanska, Henriette Mozsolits and Sheela Vemu

Instructor Guide

**Learning Objectives:**

Upon completion of each laboratory activity, the students will be able to:

* Describe the function of cellular respiration
* Identify the type of reaction of cellular respiration
* Identify the reactants and products of the cellular respiration reaction
* Interpret data sets to show the presence of CO2 as a product in cellular respiration
* Perform linear regression analysis using MS Excel 365
  + Interpret the regression output using the correlation coefficient
  + Construct a scatter plot using MS Excel 365
  + Use R-squared to judge the fit of the regression line
  + Describe the relationship between the biological variables presented in this experiment.
  + Predict the value of an output variable (response) based on the input of an input (predictor) variable.

**Background Information:**

***Biology Background Information****:*

Cell respiration refers to the process of converting the chemical energy of organic molecules into a form immediately usable by organisms. Glucose can be oxidized completely if sufficient oxygen is available.

All organisms, including plants and animals oxidize glucose for energy. Often, this energy is used to convert ADP and phosphate into ATP. There are several methods of indirectly measuring the rate of cellular respiration in organisms. One method involves monitoring changes in temperature; since the process of respiration is exergonic (produces heat). Another method is to measure either the oxygen consumption or the carbon dioxide production.

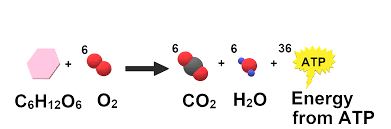
[Amoeba Sisters Cellular Respiration Video](https://www.youtube.com/watch?v=eJ9Zjc-jdys)

***Chemistry Background Information:***

In animal and plant cells, organic compounds release energy through the chemical process of cellular respiration. In this oxidation reaction, energy sources such as the sugar glucose (C6H12O6) react with oxygen (O2) to produce carbon dioxide (CO2), water (H2O) and energy (ATP):

<https://commons.wikimedia.org/wiki/User:Christinelmiller>

Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)



C6H12O6  +  6O2  →  6CO2  +  6H2O  + energy    (respiration)

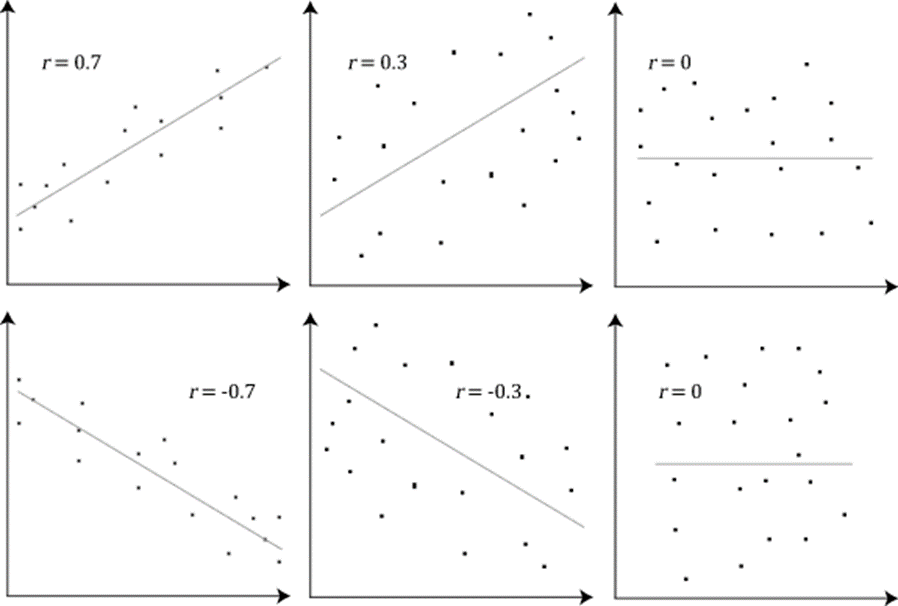
Respiration is an exothermic process, and it provides the energy to make adenosine triphosphate (ATP) that is used as a main energy source for many other cell activities.

***Mathematical Background Information****:*

Linear Regression Analysis is a statistical technique for analyzing the relationship between two variables. Regression Analysis is widely used to evaluate trends, make estimates, and for forecasting. In addition to applications in biology and chemistry, regression analysis is used in engineering, medicine, financial analysis, to name a few.

The regression model predicts the behavior of a dependent (target) variable using an independent (predictor) variable. We will use MS Excel 365 for graphing the scatter plot and performing linear regression on a data set. Excel was selected because it is available at no cost to students unlike graphing calculators. Further, proficiency in MS Excel is a useful skill for students’ future careers.

A scatterplot can be used to determine whether a linear correlation exists between two variables. Several types of correlations (r) are shown in the scatterplots below.



<https://commons.wikimedia.org/wiki/File:Pearson_Correlation_Coefficient_and_associated_scatterplots.png>

**The Lab:**

**The Biology lab:** [Cell Respiration Lab with Germinating vs Non Germinating Peas](https://collaborate.hms.harvard.edu/download/attachments/52760255/Lab3_Cellular_Respiration_S-Guide_hinton20_v4.pdf?api=v2)

*This material is based upon work supported by the National Science Foundation under Grant No.1919613. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.*

**Pre-Lab Activity with MS EXCEL 365 Instructions:**

This activity allows students to practice linear regression with sample data before the lab. The goal is to help familiarize students with the process of creating and using a linear regression before they are in the lab environment.

**Finding the Correlation Coefficient, Creating a Scatter Plot, Finding the Regression Line, and R-Squared in MS Excel 365**

A researcher wants to determine whether there is a linear relationship between a person’s age and the number of hours he or she exercises each week. The data are shown in the table below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Age, x** | 33 | 23 | 49 | 57 | 63 | 40 | 27 |
| **Hours of exercise, y** | 6 | 6.5 | 3.5 | 3 | 1 | 4 | 6 |

1. Calculate the correlation coefficient for the person’s age and the number of hours he or she exercises each week. Interpret the results in the context of the data.

*Note: The correlation coefficient indicates the type and strength of the linear relationship between two variables, x and y. The correlation coefficient is a value between -1 and 1 inclusive.*

1. Display the data in a scatter plot and use the scatter plot to describe the type of correlation.
2. Find the equation of the regression line for the person’s age and the number of hours he or she exercises each week. Add the regression line to the scatter plot. Report the R-Squared value and interpret.

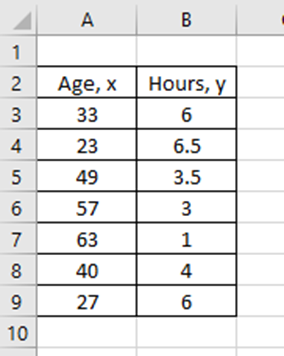
*Note: R-squared indicates how well the regression model fits the data set.*

1. Use the regression equation to predict y- values. Predict the number of hours a 45-year-old person will exercise per week.

**Instructions for Instructors with Answer key :**

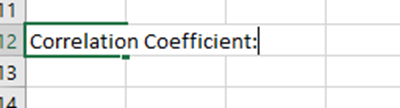
1. **To find the correlation coefficient:**

Open the Excel and enter the data in two columns.



We can find the correlation coefficient if we use the “CORREL” function in Excel.

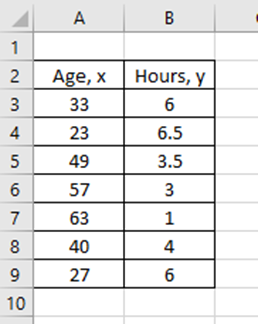
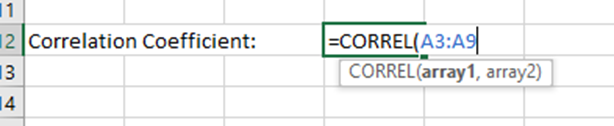
Write the phrase “Correlation Coefficient:” below the table.



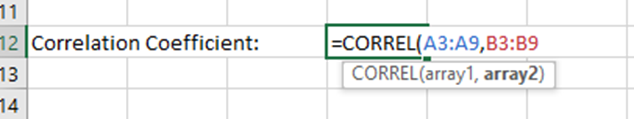
In the cell to the right of the phrase “Correlation Coefficient:” start typing **=CORREL** and you will get the function with parentheses opened that shows you how you need to input the data: ***(array 1, array 2***

Highlight the first column (which in our example is column A from A3 to A9) and when you do this, the ***array 1*** will be replaced by A3:A9

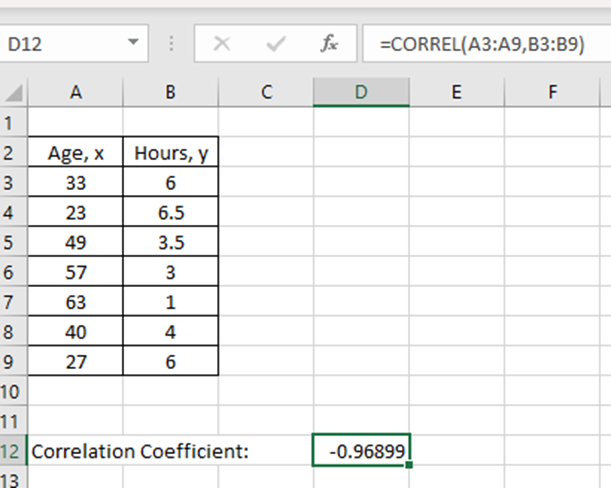
**Do not highlight the title of the column!**

After A9 enter a **comma,** then repeat the highlighting of the second column of the data:



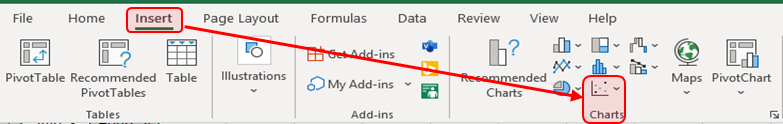
Now close the parentheses and then hit “**Enter**”, and you will get the following:



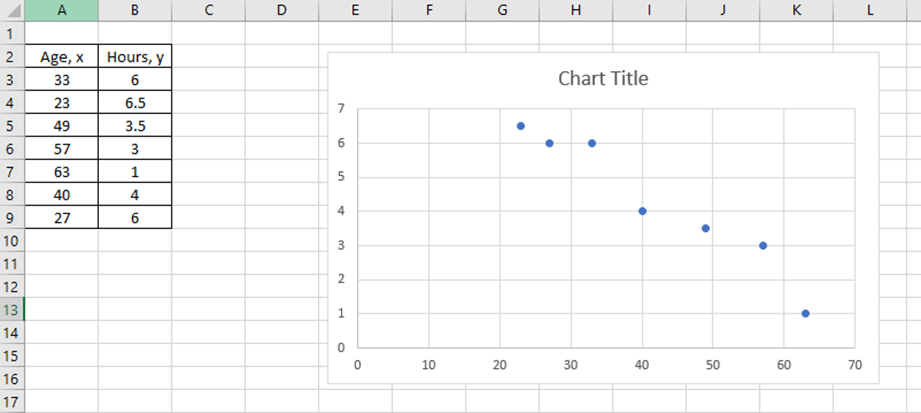
The result r = - 0.96899 suggests a strong negative linear correlation. So, as a person’s age increases, the number of hours the person exercises each week tends to decrease

1. **To create the scatter plot:**

Highlight both columns with the data (this time highlight the title of each column) then click the “Insert” tab on the top menu, then click the “Charts” tab.



Choose the first type of chart by clicking on it. This will give you the scatter plot.



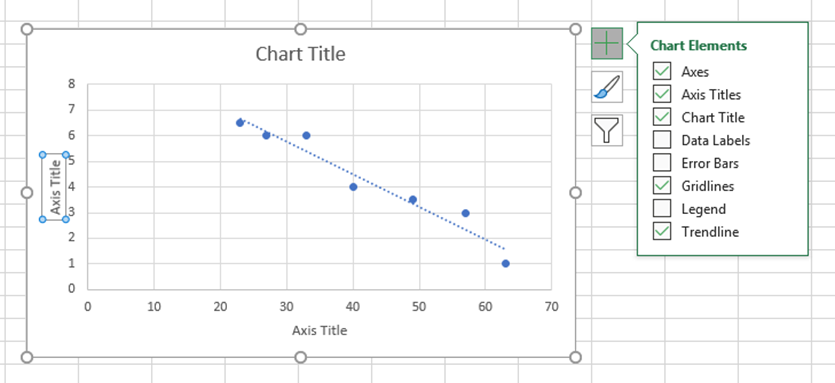
From the scatter plot, it appears that there is a negative linear correlation between the variables. As the person’s age increases, the number of hours a person exercises per week tends to decrease.

1. **To label the scatterplot, find the equation of the regression line, and the R-squared value (the coefficient of determination):**

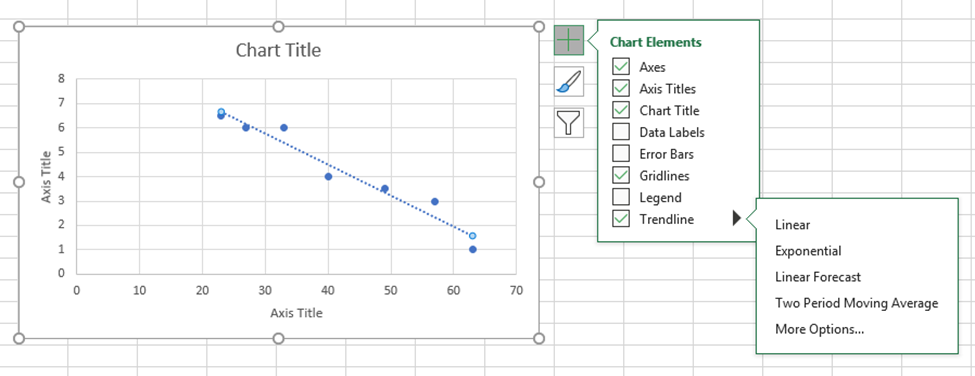
To label the scatterplot:

* Type in the Chart Title
* Titles of the axes (x- stands for the Age, y – stands for the Hours Exercised)
* Add the regression line, the equation, and the R-squared to your scatter plot

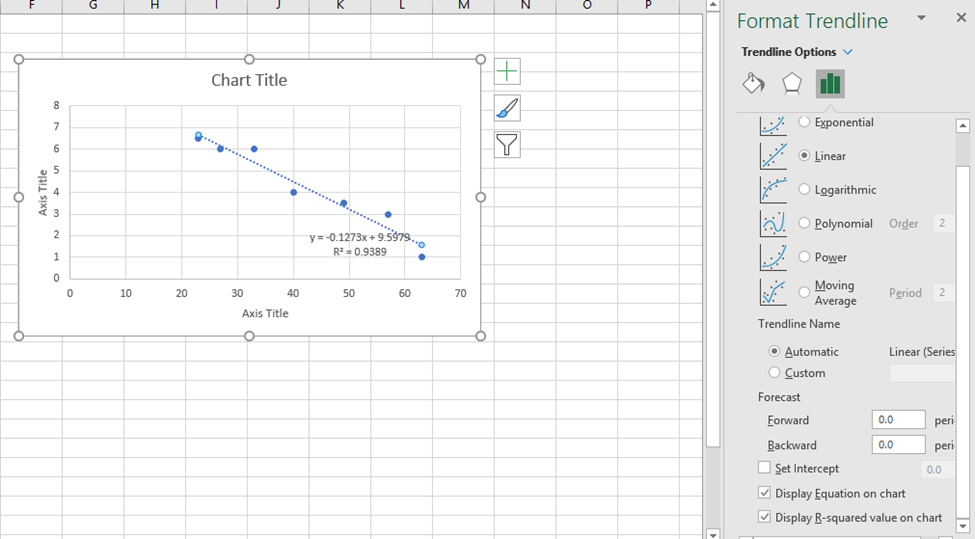
Click anywhere on the scatter plot, then click the **(+)** sign in the top right corner of the plot and check the boxes - **Axis Titles** and **Trendline**. This will also add the **linear regression line** to your scatterplot.



Next, add the equation of the regression line to the chart and the R-squared value by clicking **More Options**.

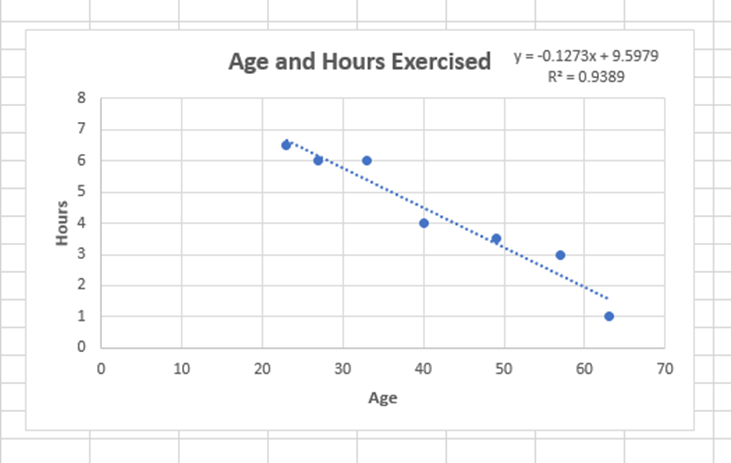


In the box to the right, check the boxes next to **Display Equation on chart** and **Display R-squared value on chart.**

****

**The linear regression equation and the R-squared will appear on the scatterplot.**

Now you can type in the chart title and titles of the axes. You can also relocate the equation of the regression line.



The regression line turns out to be *y* = - 0.1273 *x* + 9.5979

Interpretation: For each additional one unit increase in the *x* variable - age, there is a 0.1273 decrease in the y variable – hours of exercise.

The R-squared value is 0.9389

Interpretation: About 93.89% of data fit the regression line or About 93.89% of the variation in *y* - hours exercised can be explained by the relationship between *x*- age and *y-* hours exercised. The remaining 6.11% of the variation is unexplained and is due to other factors.

1. **Use the regression equation to predict y-values.**

There is significant linear correlation between the person’s age and the number of hours he or she exercises each week. So, you can use the regression equation to predict *y*-values.

To predict the expected number of hours a 45-year-old person exercises each week, substitute the age of 45 for *x* in the regression equation and solve.

*y* = - 0.1273\*45 + 9.5979 = 3.8694

A 45-year-old person is expected to exercise 3.9 hours per week.

**Additional Resources:**

* [How to run a linear regression in EXCEL](https://www.lifewire.com/how-to-run-regression-in-excel-4690640)
* [365 Simple Regression Video](https://www.youtube.com/watch?v=hr-n4v_PMmo)

**Experimental Activity (Module) with Answer Key**

This can be used in multiple modalities of instruction:

1. Additional Pre-Lab exercise for an in person lab,
2. Lab exercise for a remote lab
3. In a hybrid class, you could assign this as a pre-lab activity and collect data in the lab.

We will utilize the germinating peas to examine the process of cellular respiration. Measuring the rate of respiration can rely on the amount of oxygen taken or the amount of carbon dioxide released by the peas. There are two gases (oxygen and carbon dioxide) that are involved in cellular respiration. During cellular respiration these two gases are changing in volume.

There are many ways to measure the rate of cellular respiration in an organism. One can use a respirometer to measure the gas volume changes based on the production or consumption of the gases. One can also use a carbon dioxide gas sensor or probe to monitor the amount of gas released by the germinating peas.

**Step 1** Sample Data set given

Cellular Respiration Data – obtained with a CO2 probe

Germinating vs. Non-germinating peas

Carbon Dioxide (ppm) vs. Time

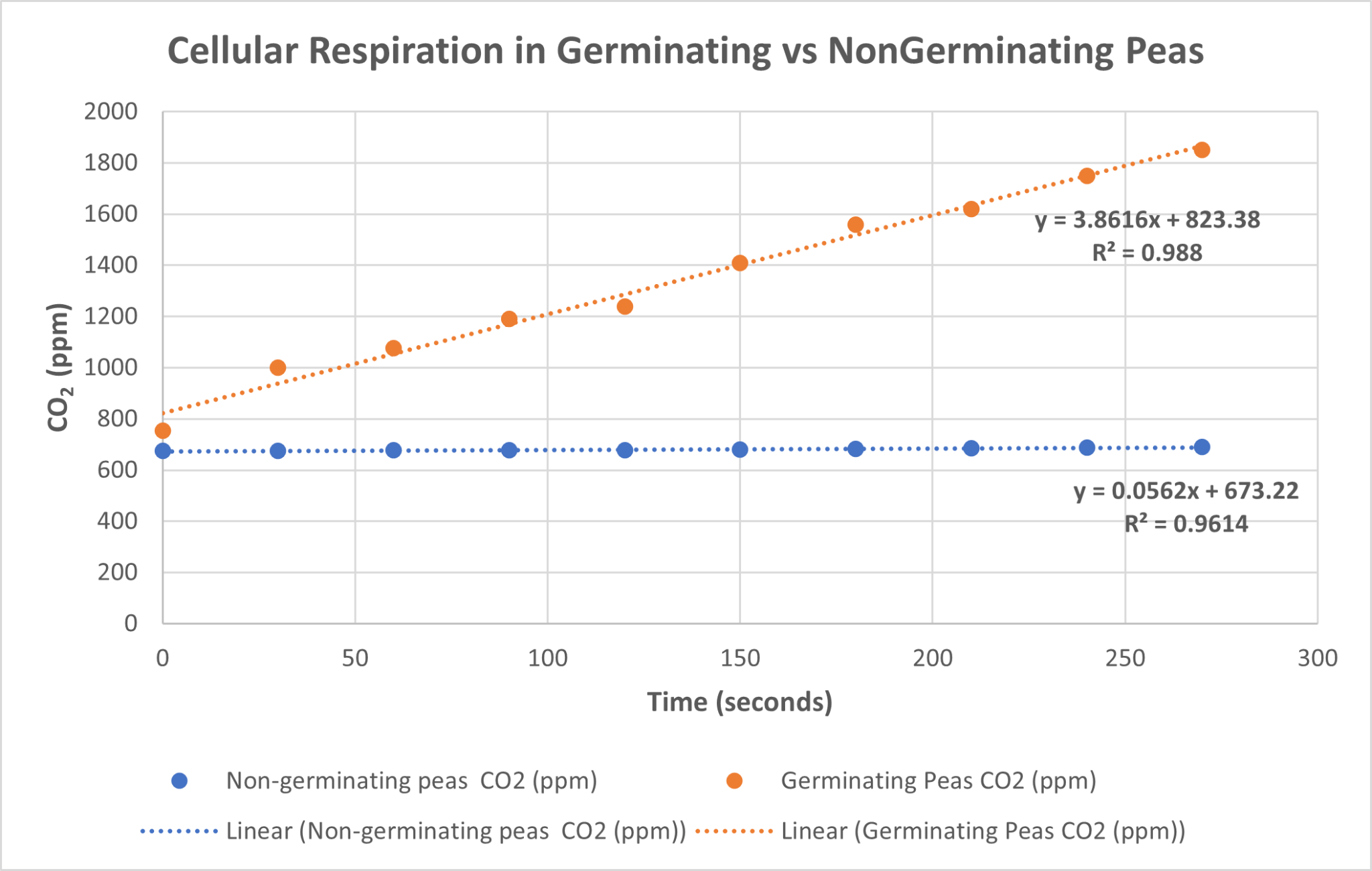
|  |  |  |
| --- | --- | --- |
| **Time**  **(seconds)** | **Non-germinating peas**  **CO2 (ppm)** | **Germinating Peas**  **CO2 (ppm)** |
| 0 | 674 | 754 |
| 30 | 676 | 999 |
| 60 | 677 | 1076 |
| 90 | 677 | 1190 |
| 120 | 679 | 1240 |
| 150 | 681 | 1410 |
| 180 | 682 | 1558 |
| 210 | 685 | 1620 |
| 240 | 687 | 1750 |
| 270 | 690 | 1850 |

**Step 2**Create a scatter plot and linear regression in MS Excel 365 using the data from Step 1

Create a linear-regression equation of the form:   **𝑦 = 𝑚x + 𝑏**

where x is time in seconds, y is CO2 concentration (ppm), m is the slope, and b is the y-intercept

**The** **Scatter Plots:**



**Step 3** Sample analysis questions with suggested answers

1. Do you have evidence that cell respiration occurred in peas? Explain.

Answer: The amount of carbon dioxide production in the chamber increases with the slope of line being positive for the germinating peas. These peas are taking in oxygen and converting their sugars into water and carbon dioxide. Look at the equation of cellular respiration to understand the importance of the reactants and products in this process

1. What is the effect of germination on the rate of cellular respiration in peas?

Answer: Since the germinating peas are sprouting, they require a more extensive amount of energy or ATP. This allows them to have high oxygen consumption rates or respiration rates in this experiment

1. Would you consider non-germinated peas “Alive”?

Answer: According to the scientific community, the peas are living because they contain an embryo that is capable of germinating to produce a new plant under optimum temperature, humidity and availability of water. If the non-germinated peas were not alive, they would not be able to germinate under the right conditions to obtain the energy required for metabolic activity,

1. Write a balanced chemical equation for cellular respiration.

Answer: C6H12O6 + 6O2 → 6CO2 + 6H2O + energy (respiration)

1. What are the reactants of cellular respiration?

Answer: Glucose and Oxygen

1. What are the products of cellular respiration?

Answer: Carbon Dioxide, Water and ATP

1. Cellular respiration is defined as \_\_\_\_\_\_ reaction.

Answer: Oxidation reaction

1. Determine the type of correlation between the variables by looking at the scatter plot. (For example, positive linear, negative linear, no linear correlation)

Answer: Both are positive linear correlations - as time passes (increases) the amount of carbon dioxide (CO2) released increases.

NonGerminating peas: R² = 0.9614

Germinating Peas: R² = 0.988

1. Compute the value of the correlation coefficient, r, for x and y. Determine the type of correlation between the variables using the correlation coefficient.

Answer: Correlation coefficient**:**

NonGerminating peas: 0.980503581 *strong positive*

Germinating peas: 0.993979156 *strong positive*

1. What is the expected CO2 level for germinating peas after 300 minutes?

Answer: Use a Linear Regression Equation to find y given x:

Germinating Peas: y = 3.8616x + 823.38

x = 300 y = ?

y = 3.8616(300) + 823.38 = 1981.86 ppm

After 300 minutes we expected the germinating peas to release about 1982ppm of carbon dioxide.